Practicum 1: Weather Statistics
January 1, 2017
Ole J. Forsberg

Problem 1: Measures of Center and Spread

This is a table of the measures of center and of spread (variation). When two measures are provided, the correct one depends on the interpretation of the histogram for symmetry. If the distribution is skewed, one should use the median (and IQR). If the distribution is symmetric, one should use the mean (and standard deviation).

From viewing the histograms, the bolded measures are most appropriate. This is because all histograms suggest the variable is significantly skewed.

Variable	Measure of Center		Measure of Spread	
Fog	**Mode**	**No Significant Fog**	**None for Categorical Variables**	
Rain	**Mode**	**No Measurable Rain**	**None for Categorical Variables**	
AvgTemp	Mean **Median**	61.46 **63**	SD **IQR**	17.48 **27**
AvgVisibility	Mean **Median**	14.87 **16**	SD **IQR**	2.17 **1**
MaxTemp	Mean **Median**	73.39 **75**	SD **IQR**	18.11 **26**
MinTemp	Mean **Median**	49.81 **52**	SD **IQR**	18.19 **31**

Interpretation

The measures of center are not particularly helpful, usually. If, however, a weather forecaster were to forget what time of year it is, then the measures of center would be his forecast. In other words, these estimates are best only if there is no additional information, like the time of year.

Problem 2: Correlation

The correlation between the minimum and maximum temperatures is r = 0.8914.

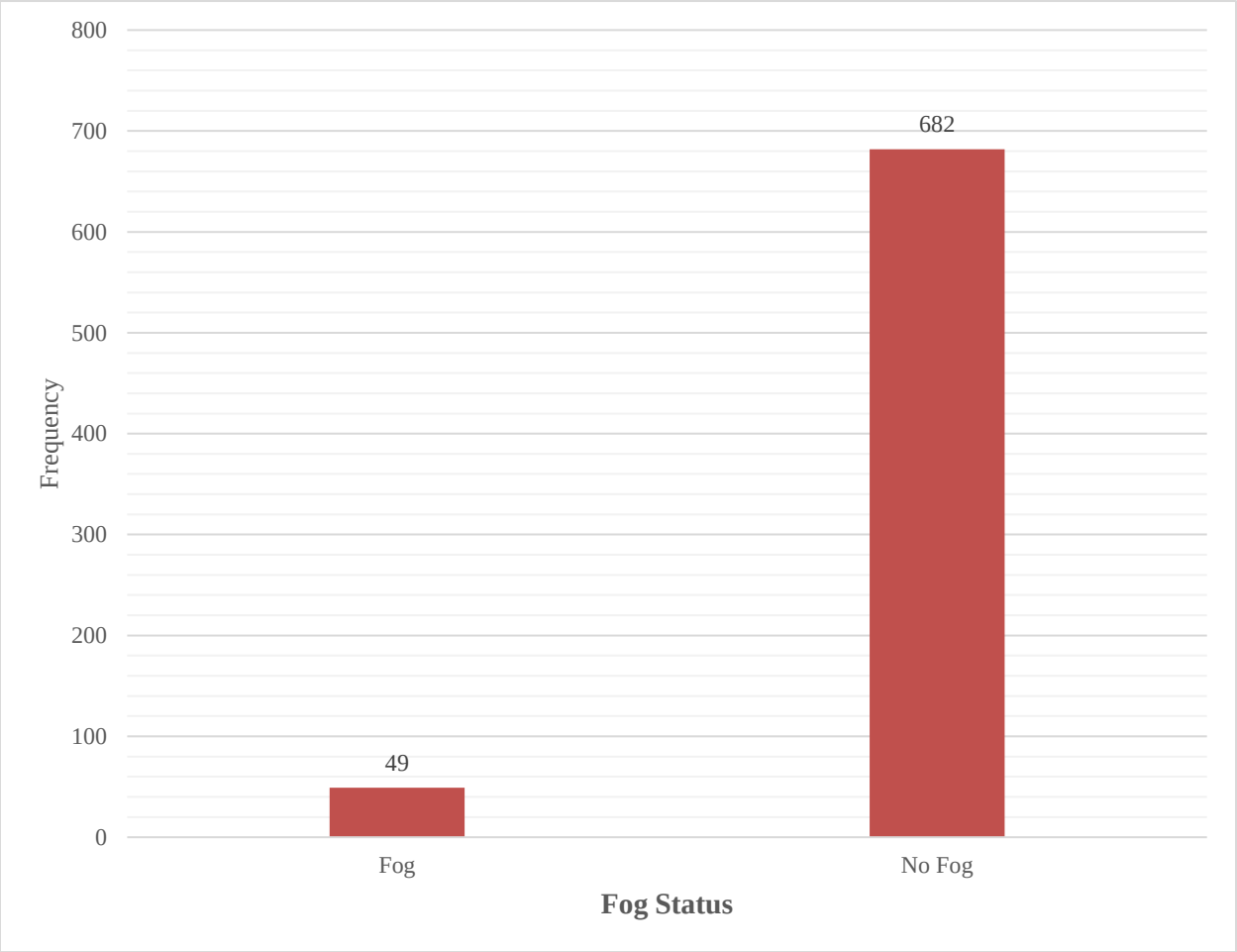
Interpretation

The correlation is a measure of linear relationship between two numeric variables. It ranges between –1 and +1. The measured correlation, r = 0.8914, is positive and close to 1. This indicates a strong linear relationship between the minimum and maximum temperatures in a day.

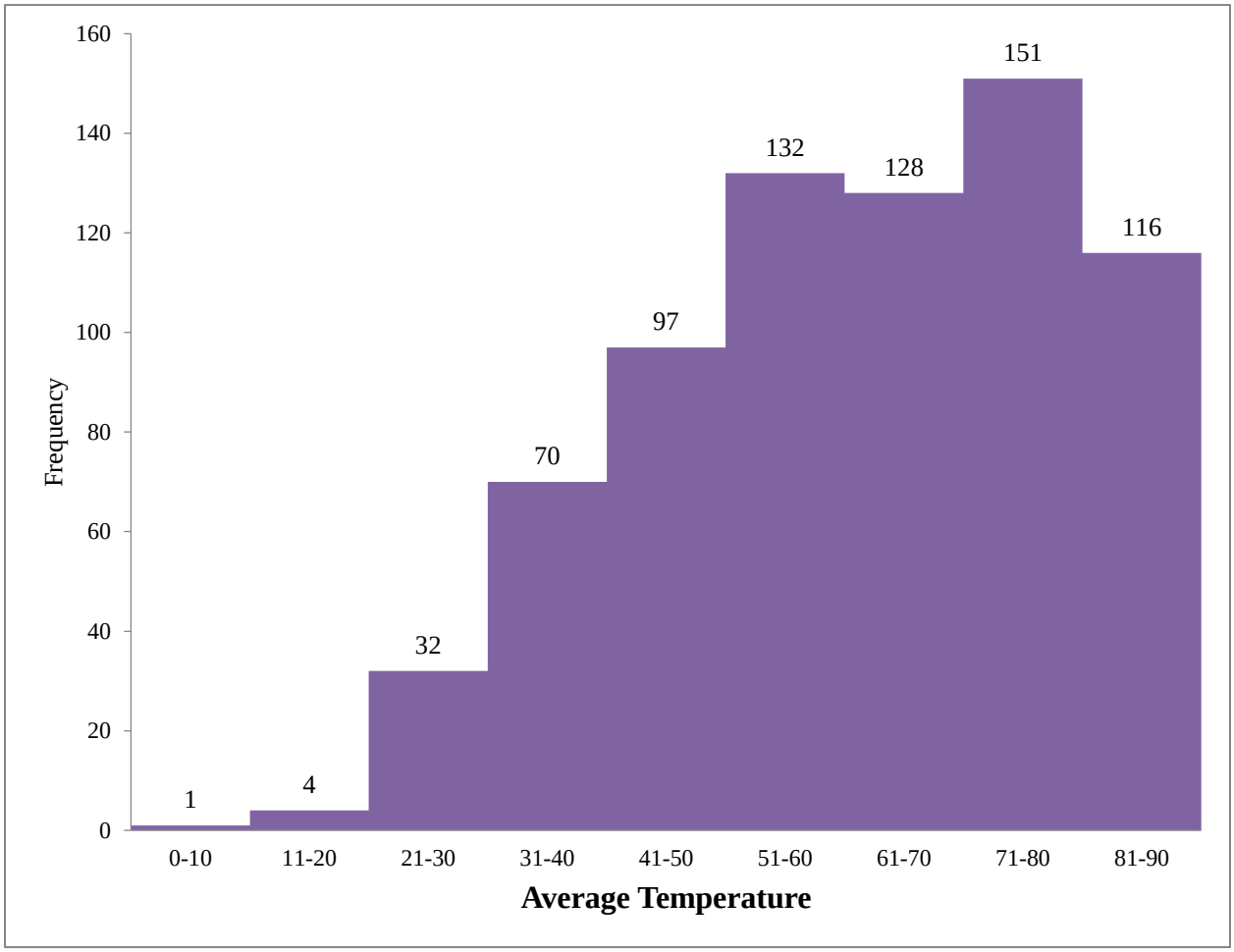
A weather forecaster, again assuming she has no other information, would predict the high temperature would be higher tomorrow if the low temperature is higher tomorrow.

Problem 3: Univariate Graphics

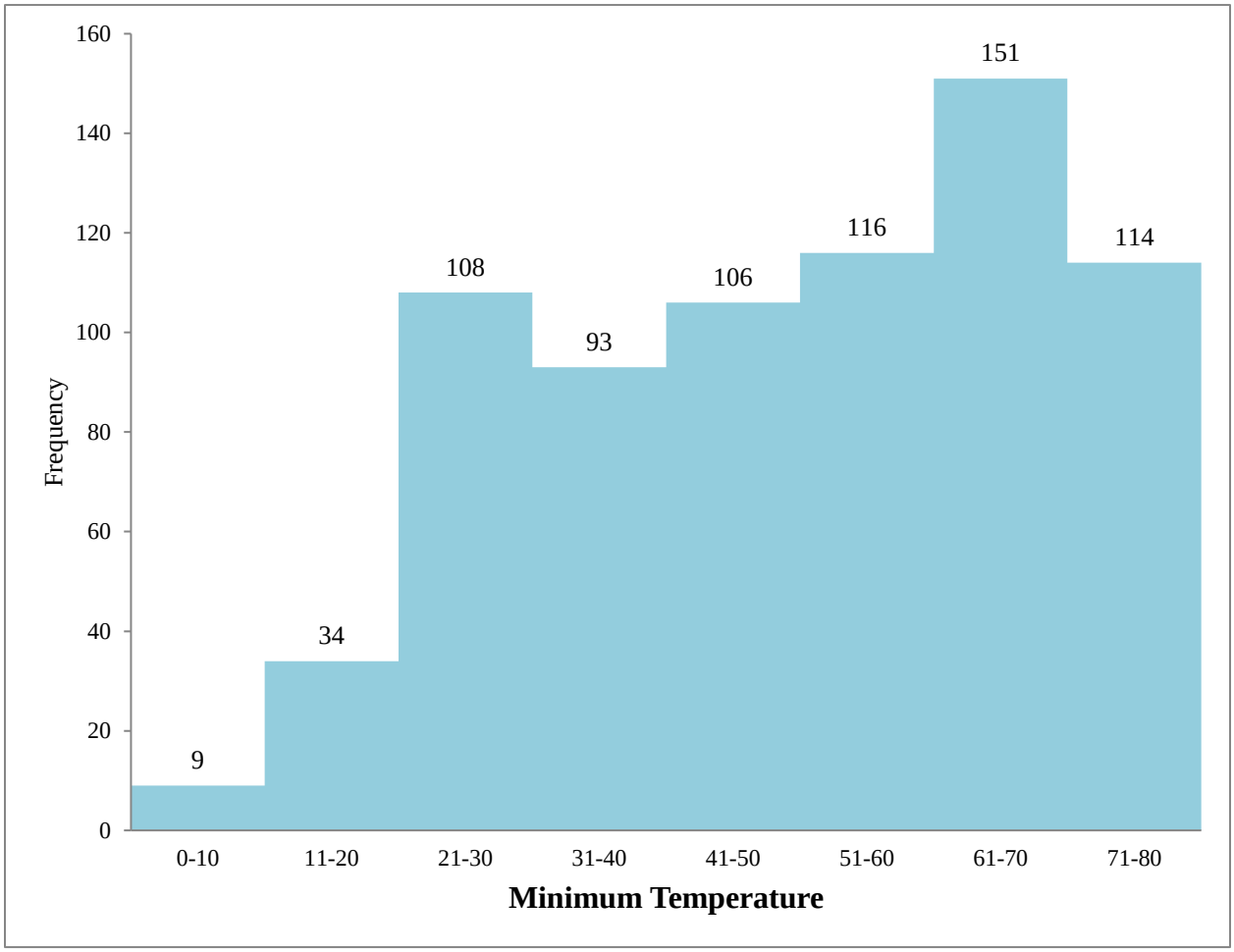
Fog Graphic



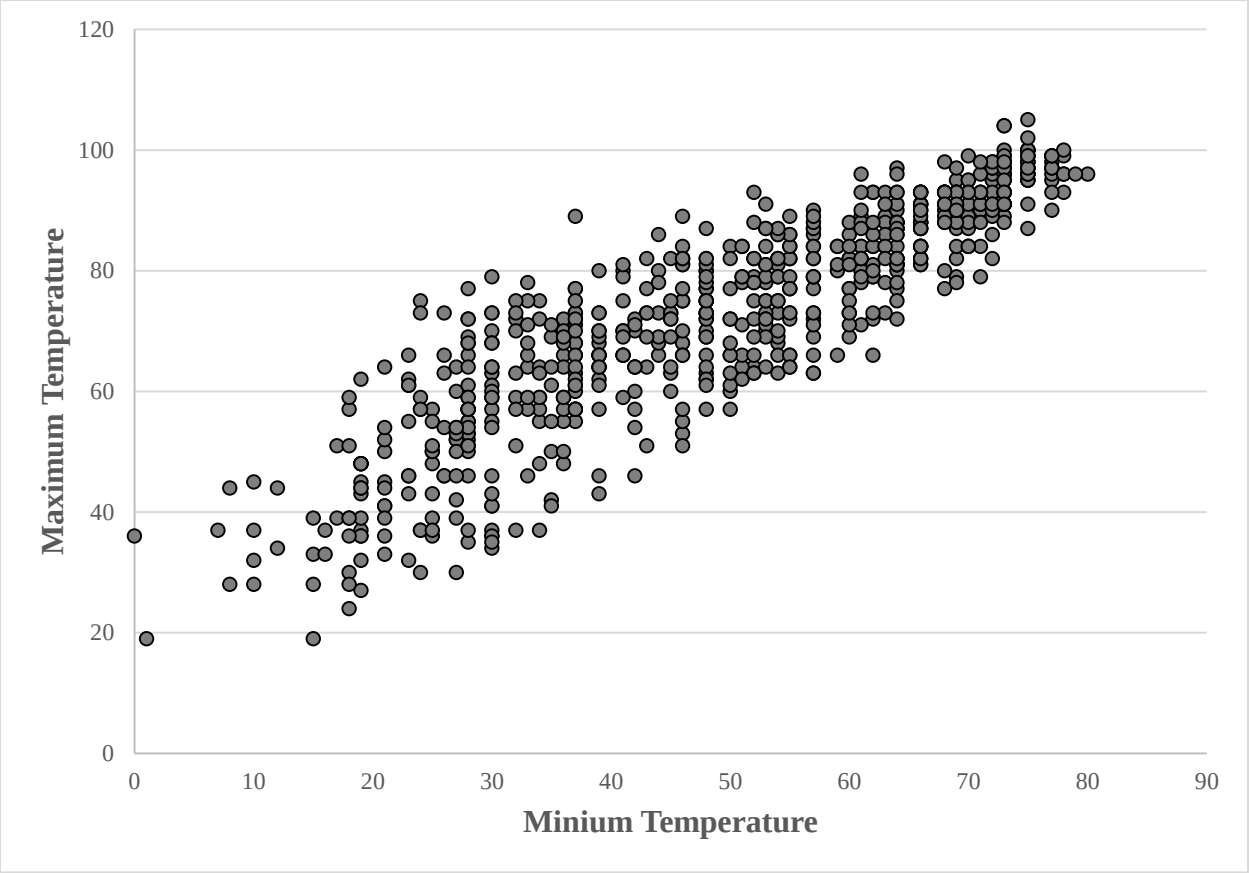
AvgTemp Graphic



MinTemp Graphic



Problem 4: Bivariate Graphic



Interpretation

Previously, we calculated the correlation between these two variables. Its value was 0.8914, which is close to +1. This scatter plot graphically shows what such a high correlation means in terms of the data.

Again, given no additional information, should tomorrow's low temperature be higher than today's, the weather forecaster would predict tomorrow's high temperature is higher, too. This is because the points seem to fall along an upward-sloping line, with random variation included.